9950-1033

DOE/JPL-956525-85/6

DRL No. 198
DRD Line Item No. SE-6

Distribution Category UC-63

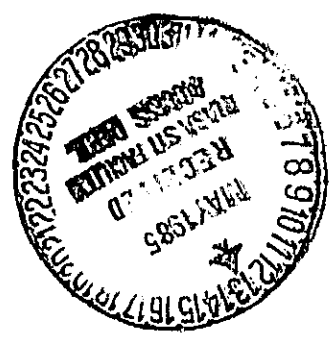
HEAVY DOPING EFFECTS IN HIGH EFFICIENCY SILICON SOLAR CELLS

Quarterly Report

For Period Covering
October 1, 1984 - December 31, 1984

By:

F. A. Lindholm and A. Neugroschel
UNIVERSITY OF FLORIDA
Department of Electrical Engineering
Gainesville, FL 32611



JPL Contract No. 956525

The JPL Low-Cost Silicon Solar Array Project is sponsored by the U.S. Department of Energy and forms part of the Solar Photovoltaic Conversion Program to initiate a major effort toward the development of low-cost solar arrays. This work was performed for the Jet Propulsion Laboratory, California Institute of Technology, by agreement between NASA and DOE.

This work was performed for the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the U. S. Department of Energy through an agreement with the National Aeronautics and Space Administration.

TABLE OF CONTENTS

Abstract.....1

I. Introduction and Executive Summary.....2

1.1 General Background.....2

1.2 Our Findings Reported Here (see Appendix for details).....3

Appendix.....6

A Model for Band-Gap Shrinkage in Semiconductors with Application to Silicon

ABSTRACT

A single mechanism is proposed as the dominant origin of energy-gap narrowing in heavily doped emitter and back-surface-field regions. The mechamism, Coulomb forces resulting in the correlated motion of majority carriers and screening by them, proposed before (by Sah), is here developed in more detail. The aim is to aid the understanding of energy-gap narrowing, and to provide a single formula for both heavily doped regions and also for regions in which holes and electrons exist together in large numbers (highly excited regions). Energy-gap narrowing influences conversion efficiency mainly through its effects on the open-circuit voltage.

I. INTRODUCTION AND EXECUTIVE SUMMARY

This report describes technical findings of work supported by contract no. 956525 for the period, October 1, 1984 to December 31, 1984. Details of the findings appear in the Appendix. In this section, we briefly summarize these findings and indicate their practical implications for solar-cell design.

1.1 General Background

One of the tasks for this contract is the identification of key parameters that describe heavily doped regions. This identification together with an understanding of the associated physical mechanisms enables the possibility of systematic design. From a performance standpoint, the heavy-doping effects reduce the open-circuit voltage severely, thus also degrading the fill factor. There is some influence also on the short-circuit current--for short wavelengths from heavily doped emitters and for longer wavelengths for a heavily doped back-surface-field region.

For the past ten years, energy-gap narrowing, commonly known in the literature as band-gap narrowing, has been put forward as one of the key parameters. It is known that the physical origin of band-gap narrowing is complicated. The detailed many-body problem has received attention from solid-state theorists. Their final results differ because of different approximations made in treating the quantum mechanical many-body problem. Moreover, their results apply strictly at T = OK, where the states relating to the impurity atoms are pure in the sense that they are not perturbed by lattice vibrations.

To further complicate the problem, many solid-state theorists assume a periodic distribution in space of the impurity atoms. This ignores the

thermal diffusion--a random-motion process--that contributes to the placement of the impurity atoms in the lattice. Thus it ignores a random component of the spatial distribution of these atoms.

On the other hand, other theorists emphasize this random component, which leads to band-edge distortion. This falls into two catagories.

*band tails

*impurity bands

The band tails have significance at higher impurty concentrations, whereas the impurity bands play a role for concentrations of the order of $10^{18}/cm^3$. Experimental evidence exists to support these claims. Some workers have suggested recently that delocalized band tails or band tails having states with relatively high mobility are largely the origin of the band-gap shrinkage.

It is highly important to identify the mechanism underlying band-gap narrowing. If one does not know the mechanism or mechanisms largely responsible, efforts to design processing to yield high conversion efficiency are frustrated.

1.2 Our Findings Reported Here (see Appendix for details)

We deal mainly with impurity concentrations above the order of $10^{18}/cm^3$. For such concentrations, experiment shows that the quantum density of states follows nearly a standard quadratic dependence. The states in an impurity band at lower concentrations move, for higher concentrations, into the majority-carrier band. Band tails exist, but we have evidence to believe

that they influence mostly transport of minority carriers and have only a slight influence on the energy gap for concentrations of the order of $10^{19}/cm^3$.

Our findings appear in detail in the Appendix, a manuscript written by P.T. Landsberg, A. Neugroschel, F.A. Lindholm and C.T. Sah. A brief overview follows.

A single mechanism is proposed as the dominant origin of energy-gap narrowing in heavily doped and highly excited regions of silicon devices (and also of devices made of other semiconductors). The mechanism involves the correlated motion of the electrons and holes resulting from Coulomb forces. The resultant (Debye) screening reduces the energy of the system, thereby reducing the energy gap. In the literature, Sah first proposed this as the one mechanism that could produce true energy gap narrowing--that is, energy gap changes not arising from band edge distortion (bandtails and impurity bands). Our work fills in the details of Sah's proposal and tries to correct earlier attempts by others to develop a simple model involving screening.

In addition to the development, a survey of earlier experiments, and of new ones obtained by us, is made. The agreement between the model and experiments on n- and p-type silicon is good as far as transport measurements near T = 300K are concerned. It is good in the sense that the discrepancies between theory and experiment are no worse than the discrepancies between the experimental results of various authors. It also gives a good account of recent optical (photoluminesence) determinations of energy-gap shrinkage at 5K.

The most general form of the result of energy-gap shrinkage appears in Eq. (21) of the Appendix. For the degenerate limit, the model gives, in gaussian units

$$\Delta E_G = 126.6(m^*/m)^{1/2}\ (\varepsilon/10)^{-2}(n/10^{18})^{1/6} \qquad (3)$$

In regard to transport derived measurements of the energy-gap shrinkage, the model agrees best with those interpretations that do not assume equal minority and majority carrier mobilities. Elaboration on this remark appeared in the preceding Quarterly Report to JPL.

A MODEL FOR BAND-GAP SHRINKAGE IN SEMICONDUCTORS WITH APPLICATION TO SILICON

by

P. T. Landsberg[1,2], A. Neugroschel[1], F. A. Lindholm[1] and C. T. Sah[3]

[1]Department of Electrical Engineering
University of Florida
Gainesville, FL 32611

[2]Permanent address: Department of Mathematics
University of Southampton, United Kingdom

[3]Department of Electrical Engineering
University of Illinois
Urbana, IL 61801

ABSTRACT

A model for bandgap shrinkage in semiconductors is developed and applied to silicon. A survey of earlier experiments, and of new ones obtained by the authors, give an agreement between the model and experiments on n- and p-type silicon which is good as far as transport measurements in ~300 K range by various authors are concerned. In fact, the discrepancies between theory and experiment are no worse than the discrepancies between the experimental results of various authors. It also gives a good account of recent optical determinations of band-gap shrinkage at 5 K.

The model is based on Debye screening and, apart from effective masses, temperature, dielectric constant and carrier concentrations, which must enter any theory of gap shrinkage in some form, it is parameter-free. The model gives in gaussian units $\Delta E_G = 126.6(m^*/m)^{1/2}(\varepsilon/10)^{-3/2}(n/10^{18})^{1/6}$ meV in the degenerate limit.

1. INTRODUCTION

There has been much discussion of the problem of band-gap shrinkage in the recent literature. The problem is of considerable importance because the band-gap is one of the key parameters in semiconductor devices. Heavy doping, which leads to the effect, has been used increasingly as part of the drive towards microminiaturization. However, key aspects of the subject are still controversial. "Discrepancies in the values extracted from experiments" have been noted [1]. At the theoretical level there are also serious discrepancies which arise from the intrinsic complexities of many-body calculations and the difficulty one has in estimating errors. Some contributions to the band-gap shrinkage are not agreed upon as regards sign, while other contributions are estimated by some authors as large while others neglect them. We refer to important recent papers [1-3] for more details and for developments of the theory. Some time will clearly elapse before these matters are clarified, and this justifies a search for other approaches which are less ambitious but may capture some aspect of the essence of the problem, and lead to an adequate estimate. We pursue one such line here.

Three points are to be emphasized at the outset.

(i) Simple Useful Formula

Our main result for the band gap shrinkage ΔE_G is equation (1), below. We observe that this formula represents the experimental facts to good approximation both at 5K (see Fig. 4) and at 300K (see Fig. 3). Thus the formula given can at the very least be regarded as a semi-empirical summary of experimental results. It may be expected to cover the experiments for the whole temperature range 5K - 300K and beyond. Section 4 discusses the comparison with experiment. The formula (1) applies to heavily doped

semiconductors and to highly excited regions of semiconductors such as occur either under strong irradiation or in the transition region of p-n junction subjected to a large forward voltage. The formula thus meets the need of many experimentalists and device engineers who require a simple analytical expression which can be introduced into other considerations that are sensitive to gap shrinkage. The device work is normally conducted at room temperature (while the many-body theories are worked out for T=0).

(ii) Jellium Model

What is the theoretical status of equation (1)? For its derivation the charged ions of the solid both of the host atoms and of the impurity atoms are imagined uniformly smeared out within the volume of the solid to supply a uniform background so that both hole and electron gases are electrically neutral. This "jellium" model of two gases of interacting electrons or holes is of course well-known. But also allowed to be present are neutral atoms which, on being ionized, can give rise to electron-hole pairs. The long-range part of the Coulomb interactions leads to complicated and correlated motions among the particles of the two gases (which one may consider separately). The n-particle wavefunctions for these systems are not known. However, it is reasonable to assume that these correlations leave the short-range part of the interactions more or less intact, thus leading to short-range or screened potentials acting between largely independent particles (the quasi-particles). The screening parameter (equation (21), below) takes account of this situation. In such a model, interactions with ions, band structure, location of impurities, inter-valley scattering, etc. are clearly neglected. Even in T=0 many-body theories, however, these are problems whose importance has not yet been ascertained with certainty. Thus inter-valley scattering was

emphasized in [1], but largely ignored in [4], while the magnitude of the effect of impurity scattering on band gap shrinkage is also not certain [2,4]. Other unsettled questions of current research which arise in these treatments include questions of linear versus non-linear screening [3] or ordered versus disordered impurities on a lattice [1-3]. By ignoring these questions here we merely follow through the jellium (or screening) model.

Our procedure has great simplicity, and it is also highly approximate, and so does not compete with the much more sophisticated many-body theories. These are of necessity limited to T=0K at the present time, and so do not yield as yet what is needed by experimentalists. It is crucial to observe that we do not ignore these recent advances, [1-3]. Instead, we seek a direct and intuitive route to a formula which, while approximate, is likely to be useful. In fact, our paper represents a challenge, in that equation (1) should be deducible from many-body theories by appropriate approximations.

In our derivation we go back to the Debye-Hückel theory of electrolytes, as expounded by W. J. Moore [5] and as used in a related context by Morin and Maita [6] in a standard paper. Some generalizations of this early work (section 2) enable one to give the simple derivation required (section 3).

(iii) Earlier Related Work

Equation (1) has appeared before. Inkson [7] obtains it as his component $\Delta E^{(1)}$ and in the review [8] it appears similarly as equation (3.13). Equation (1) was also used by Sah and collaborators [9] to discuss experimental results. On the other hand similarly simple formulae, but different in essential details, appear elsewhere [10-12]. In the present paper it is however, given a central status and it is obtained by a simple derivation for the first time.

Our formula yields a larger value for ΔE_G for degenerate material than does the work [12]. The factor is of order 1.33 and can be quite significant: A typical difference is 200 meV compared with 160 meV. The difference in the coefficients is due to the following circumstance. Both here and in [12] the gap narrowing is attributed to a drop in an energy, X say, in the presence of screening by carriers, as against the energy X in the absence of screening. The energy X used for this purpose in [12] is the energy stored in the field. As it is not obvious that this is the best choice, we chose in this paper for X the energy needed to create and separate an additional electron-hole pair without imparting kinetic energy to either particle. This seems to correspond rather directly to the energy gap concept. The results differ therefore because the models differ in an important point of detail.

2. THE MORIN-MAITA FORMULA

We commence with a discussion of the formula to be used for band-gap shrinkage. It is

$$\Delta E_G = e^2\lambda/\varepsilon \quad , \tag{1}$$

where e is the numerical value of the electron charge, ε is an appropriate dielectric constant of the material and

$$\lambda = \tfrac{1}{2}\lambda_1 \quad [6] \tag{2}$$

or

$$\lambda = \lambda_1 \quad [9] \quad . \tag{3}$$

where

$$\lambda_1^2 \equiv 4\pi e^2(n + p)/\varepsilon kT \quad . \tag{4}$$

The quantity (3) is essentially the Debye screening parameter. In [9] the MKS (S.I.) system was used, which has here been changed to c.g.s. (Gaussian) system by replacing the dielectric constant relative to vacuum by $\varepsilon/4\pi$. The discrepancy between (2) and (3) will now be explained.

Suppose the electrostatic potential due to a charge e can be written

$$\phi(e,r) = b(r)e \quad , \qquad -$$

where b is typically $1/\varepsilon r$, but can have other values (see equation (10), below). The electrostatic energy due to another charge e' at the appropriate point in this field is clearly bee', or for e' = e,

$$U(e,r) \equiv e\phi(e,r) = b(r)e^2 \quad . \qquad (5)$$

This is all that is needed to obtain the energy of the charge e in the potential $\phi(e,r)$. The connection between U and ΔE_G is made in equation (14).

In [6] the energy calculation is approached by estimating the work done in charging an ion reversibly in its appropriate position r up to its total charge e. This is taken to be

$$V(e,r) \equiv \int_{q=0}^{e} q\, d\phi(q,r) = \frac{1}{b(r)} \int_{q=0}^{e} \phi(q,r)d\phi(q,r) = \frac{[\phi(e,r)]^2}{2b(r)} = \frac{U(e,r)}{2} \quad . \qquad (6)$$

It was overlooked that this answer is half that which one would expect since it is assumed that the potential b(r)e also starts from zero, thus lowering the energy V below what it should be.

In Moore's treatment

$$b(r) = \lambda/\varepsilon \quad \text{(independent of } r\text{)} \quad , \tag{7}$$

and it is obtained by splitting the normal Coulomb potential off the screened Coulomb potential:

$$\frac{e}{\varepsilon r} \exp(-\lambda r) \approx \frac{e}{\varepsilon r} - \frac{e\lambda}{\varepsilon} \quad . \tag{8}$$

The first term is the potential at distance r due to a charge e in a medium of dielectric constant ε. The second term is the potential due to the other ions in the electrolyte or semiconductor. The energy (5) or (6) (depending on what one wants to calculate) then gives a reduction which takes into account the electrostatic interaction among the current carriers. In [6] the combination of equations (2), (5) and (7) was taken to lead to the band-gap shrinkage (1). This was done by simply citing reference [5] where the second term in equation (8) was inserted in a free energy. Morin and Maita interpreted this as a change in chemical potential and hence in energy gap, and so applied the result as a correction to the np-product. The job was done in nine lines of text.

It is useful, in order to bring out the semiconductor physics involved more explicitly, to indicate first a generalization of the above argument. One need not expand the exponent as in (8). It is perfectly valid to replace (8) by the exact result

$$(e/\varepsilon r) \exp(-\lambda r) = e/\varepsilon r - (e/\varepsilon r)[1 - \exp(-\lambda r)] \quad . \tag{9}$$

The second term of (9) has the same interpretation as the second term in (8). Thus (7) is replaced by

$$b(r) = [1 - \exp(-\lambda r)]/\varepsilon r \quad . \tag{10}$$

The generalized and corrected [for the factor 1/2 in (2)] Morin-Maita band-gap shrinkage is, using (10) in (5),

$$U(e,r) = (e^2/\varepsilon r)[1 - \exp(-\lambda r)] \quad . \tag{11}$$

It depends on the distance r from the given charge for which the band shrinkage calculation is made.

The introduction of concepts from electrochemistry [13] is not helpful for the present purposes, as it tends to distract from the essentials of the argument.

3. THE PRESENT MODEL

We continue to focus our attention on formula (1), with a view to arriving at it by a direct path suitable for semiconductor work. The first step now is to create in the jellium an electron-hole pair which is in a bound state for a very short time. The distance, a say, between the maxima of their wave packets is of the order of an Angstrom, so that the effect of the smeared out electron and hole densities will not affect the energy, W say, to create the pair. The normal Coulomb potential acts between the particles for $r > a$ and is cut off at $r = a$. Such cut-offs are often needed for small r as the Coulomb potential diverges as $r \to 0$. Imagine now the hole to be fixed at its instantaneous position, $r = 0$ say, and the electron to be removed to infinity

against the Coulomb attraction starting at the cut-off distance. (Alternatively one may regard the origin of the coordinate system to move with the hole.) As remarked in section 1, the jellium consists of quasi-particles with short-range, or screened, potentials acting between them. The screening parameter (of equation (3)) is approximated as a constant (it really depends on the wave vectors involved in the Coulombic collisions). With these approximations the total energy supplied to create the pair, and to separate it, is (Fig. 1)

$$E_G(n,p) \equiv W + (e^2/\varepsilon a)\exp(-\lambda a) \quad . \tag{12}$$

This quantity is interpreted as the minimum energy between a hole in the valence and an electron in the conduction band, where n,p are the carrier concentrations. We thus exlude the Burstein-Moss shift and use E_G in the sense of $E_{gap,2}$ of Mahan [3]. If the semiconductor is highly nondegenerate, then screening can be neglected and (12) yields

$$E_G(0,0) = W + (e^2/\varepsilon a) \quad . \tag{13}$$

As already explained, W and a are to be approximated as concentration-independent. By subtraction,

$$\Delta E_G \equiv E_G(0,0) - E_G(n,p) = (e^2/\varepsilon a)[1 - \exp(-\lambda a)] \; [= U(e,a)] \quad . \tag{14}$$

Thus one does indeed arrive at expression (11), but it is now rather more firmly linked to semiconductor concepts, a link having been established in the form $\Delta E_G = U(e,a)$. Using Debye or Thomas-Fermi screening gives the same result in the limit of extreme degeneracy, namely for n^+ material

$$\lambda^2 = \left(\frac{3}{\pi}\right)^{1/3} \frac{4e^2 m_n}{\varepsilon \hbar^2} n^{1/3} \quad , \text{ i.e.} \tag{15}$$

$$\lambda = \left(\frac{m_n/m}{\varepsilon/10}\right)^{1/2} \left[\frac{n}{10^{18}}\right]^{1/6} \times 0.0863 \text{ A}^{-1} \left[\approx 0.08\left(\frac{n}{10^{18}}\right)^{1/6} \text{ A}^{-1} \text{for Si}\right] \tag{16}$$

with $\varepsilon = 11.7$ and for $m_n = m$, where m_n is the density-of-states effective mass for electrons.

The approximation

$$\exp(-\lambda a) \sim 1 - \lambda a \tag{17}$$

requires the constraint $\lambda a \lesssim z$ where $z \sim 1/4$ (see Appendix). Hence from (16)

$$a \lesssim \left(\frac{\varepsilon/10}{m_n/m}\right)^{1/2} \left(\frac{10^{18}}{n}\right)^{1/6} 11.59z \text{ A} \left[= 12.53z\left[\frac{10^{18}}{n}\right]^{1/6} \text{ A for Si.}\right]$$

For $n \sim 10^{20}$ cm^{-3} in Si, one has $\lambda = 0.17 A^{-1}$ and $a \lesssim 1.45$ A if $z \sim 1/4$A. Adopting (17), one finally obtains equation (1) from equation (14).

Numerically, (1) and (15) yield

$$\Delta E_G = 126.6 \left[\frac{m_n/m}{(\varepsilon/10)^3}\right]^{1/2} \left[\frac{n}{10^{18}}\right]^{1/6} \text{ meV} \tag{18}$$

where n is in cm^{-3}. For silicon with $\varepsilon \simeq 11.7$ and $m_n \simeq m$, this gives $\Delta E_G \sim 215$ meV at $n \sim 10^{20}$ cm^{-3}. In fact, one can put, using $\varepsilon = 11.7$ for silicon,

$$\Delta E_G = 215(m_n/m)^{1/2}\left[\frac{n}{10^{20}}\right]^{1/6} \text{ meV} \quad . \tag{19}$$

As noted in section 1, this is 1.33 times larger than a similar formula given by Lanyon and Tuft [12] for the degenerate limit.

Although the band-gap shrinkage has been determined in (14) in terms of $\underline{a}$ and λ, a more sophisticated theory is needed to estimate $\underline{a}$. The importance of $\underline{a}$ arises from the fact that it determines the relative contributions of the two terms in (12) and (13). The beauty of the present treatment is that no commitment needs to be made as regards the numerical value of $\underline{a}$. The reason is that we need only (14) from which $\underline{a}$ cancels if (17) holds, so that the actual value of $\underline{a}$, which enters only (12) and (13), is not required.

The ratio of the two terms in (13) is

$$R \equiv \frac{e^2/\varepsilon a}{W} = \left[\frac{\varepsilon a E_G(0,0)}{e^2} - 1\right]^{-1} \quad . \tag{20}$$

A divergence in R occurs when $\underline{a}$ is so small that $W = 0$. So $\underline{a}$ must be large enough to keep R finite. For $n \sim 10^{20}$ cm^{-3} the limiting value of $\underline{a}$ is given by $\underline{a} \gtrsim 1.2$ Å for silicon. On the other hand, it must by virtue of (17) be small enough for $\lambda a < z$, i.e. R must lie above the minimum value determined by $z/\lambda \propto n^{1/6}$. Thus one can draw a curve such that, for given n, all finite R-values lying above this curve are permitted and all those lying below the curve are forbidden. This curve is shown in Fig. 2 for Si and for $z = 1/4$. It shows that the theory holds for a very wide range of possibilities as regards the relative contributions of the two terms in (13). For $n \sim 10^{20}$ cm^{-3} $\underline{a}$ must be less than about 1.5 Å.

The fact that the two quoted values of $\underline{a}$ lie well within the interatomic distance in silicon (5.42 Å) makes the model internally consistent in the

sense that one does not expect screening to be of importance for distances smaller than the interatomic distance.

The ratio of the two terms in (12) is

$$R' \equiv \frac{(e^2/\varepsilon a)\exp(-\lambda a)}{W} < R \quad .$$

Hence if R can lie above a certain value for given electron concentration n, then the freedom for R' is even greater in that it can assume any value above a smaller number.

In the Appendix it is shown that the theory begins to fail unless the electron concentration lies below a critical value which is

$$n_c \sim 4 \times 10^{20} \text{ cm}^{-3}$$

in the case of Si.

The general screening length must make allowance for electrons and holes and has many interconnections with other parts of physics [14]. One finds for a parabolic band

$$\Delta E_G = \frac{e^2}{\varepsilon}\left\{\frac{4\pi e^2}{\varepsilon kT}\left[N_c F_{-1/2}(\gamma_c) + N_v F_{-1/2}(\gamma_v)\right]\right\}^{1/2} \qquad (21)$$

where $[F_{-1/2}(a) \equiv (1/\sqrt{\pi})\int_0^\infty \{x^{-1/2}/[1 + \exp(x-a)]dx\}$ and $\gamma_c \equiv (F - E_c)/kT, \gamma_v \equiv (E_v - F)/kT$ and F is the Fermi level. It is this formula which will be adopted here for all concentrations of holes and electrons.

4. COMPARISON WITH EXPERIMENT

Energy-gap shrinkage ΔE_G obtained from our previous transport measurements on Si:As n-type layers at ~ 340 K [15] are shown in Fig. 3 as

full circles. A new result, obtained here, for Si:B p-type layers from transport measurements is shown as a full triangle. This new result was obtained using a p^+-n-p transistor with uniformly doped ($N_A \approx 5 \times 10^{19}$ cm^{-3}) 0.2 μm thick emitter. The measurement method was the same as in [15]. The data was obtained in the range 320-380 K. Assuming ΔE_G to be fairly temperature-independent in this range, it seemed reasonable to associate it with 340 K. The curve is based on our model, equation (21), and gives good agreement for $m_n/m = m_p/m = 1.1$ and a majority carrier concentration n or p $\gtrsim 5 \times 10^{19}$ cm^{-3}.

For lower concentrations our model underestimates ΔE_G. This is to be expected since the impurity band or tail electrons contribute to the screening, so that our estimate of λ in (21) is expected to be an underestimate. At the concentrations well above 10^{19} cm^{-3} the impurity band is absorbed in the conduction band and one can return to a parabolic band. This sudden change near the metal-insulator transition has been noticed for many other physical parameters [16].

In much of the literature Boltzmann statistics has been used to infer ΔE_G even if the semiconductor was degenerate. These data have been recalculated so as to isolate band-gap shrinkage and remove the effect of degeneracy.

The results are shown in Fig. 3 and compared with equation (21). The experimental points are rather dispersed, but the general agreement is quite good. One reason for the deviations may be that some of the transport data in Fig. 3 were obtained from regions with nonuniform doping and the values of ΔE_G reported are for the average doping concentration which is generally different from ΔE_G for the uniformly doped samples. The reason for this is that non-uniform doping gives rise to quasi-electric fields which leads to additional drift. In some other cases errors may have been introduced in the

evaluation of ΔE_G by the use of geometrical corrections to the one-dimensional model.

The agreement between the model and experiment approaches the agreement between different experiments, and so can be regarded as satisfatory. It is marginally better for the high concentrations ($> 5 \times 10^{19}$ cm^{-3}) which are of the greatest practical importance for the heavily-doped regions of the emitters in bipolar transistors, in solar cells and in other semiconductor devices.

The experimental estimate of ΔE_G by other workers tends to be somewhat smaller than ours for concentrations larger than about 5×10^{19} cm^{-3}. The reasons for this difference were discussed in detail elsewhere [23,24], but they are briefly repeated here. The ΔE_G comes from the measured minority-carrier current which is a function of a product $\mu[\exp(\Delta E_G/kT)]$, where μ is a minority-carrier mobility. Now if one uses a conventional assumption that the minority and majority carrier mobilities are identical, then one obtains ΔE_G as reported by others and shown in Fig. 3. In contrast our analysis [15] assumes only a temperature independence of μ in the measurement range, but it does not assume the magnitude of μ, which is not well known [8,23,24]. Thus our interpretation of the transport data is more physical. The agreement between our transport data for the degenerate concentration range and the theory is excellent.

Wagner's recent 5 K photoluminescence and photoluminescence excitation data [25] are shown in Fig. 4. N- and p-type material appear to lie on the same curve of bandgap shrinkage against concentration. This suggests that at this temperature the optical characteristics are insensitive to this difference. We have therefore compared experiment with three theoretical curves based on (21), using the non-committal symbol m* for electron and hole

effective mass. The actual values for heavily-doped silicon are not well known [16,26], but those for intrinsic silicon at 5 K are $m_n/m = 1.06$ and $m_p/m = 0.59$ [26]. These values have been included in the figure, though the best value for agreement with experiment is $m^*/m = 0.45$.

Some 2.4 K photoluminescence experiments have recently been used to infer the bandgap shrinkage appropriate to 300 K [27,28]. This data, corrected for Fermi-Dirac statistics, is given in Fig. 5. For intrinsic silicon at 300 K the effective masses of electrons and holes are $m^*/m \sim 1.1$ [9,15,29]. Although the effective masses for heavily-doped silicon at this temperature are again not well-known, the model is seen to give satisfactory agreement for $m^*/m = 1.1$. The transport data (Fig. 3) also shows good agreement between theory and experiment at ~340 K for $m^*/m \sim 1.1$. The optical absorption experiments [30-32] have not been included in Figs. 5 and 6 as it is suggested that they underestimate ΔE_G [25].

Lastly we wish to point out that the band-gap shrinkage in doped material has important contributions from one band only: the screening parameter λ depends then on one concentration. For high optical excitation, however, both electrons and holes contribute to the screening parameters, which is thus substantially increased (see equation (21)). This results in a larger band-gap shrinkage, if the effects of any impurity-bands or tail states are neglected. This is illustrated in Fig. 6 which predicts the order of magnitude of expected experimental results. Figure 6 also shows the theoretical ΔE_G for n-type GaAs. This is notably smaller than for Si since $(m_n)_{GaAs}/(m_n)_{Si} \sim 0.066/1.1$. The relevant experimental data is based on optical absorption (for a review see [8]) and is therefore not expected to be reliable. It is not shown in Fig. 6.

5. CONCLUSION

A model of band-gap shrinkage in semiconductors has been developed. Its application to n- or p-type silicon gives good agreement with both transport and optical measurements at 5 K and in the range 320-380 K. The discrepancies between theory and experiment are on the whole no greater than the discrepancies inherent in the experimental determinations. Our model would be expected to underestimate ΔE_G in the non-degenerate regime. Here the impurity-band and tail states are not merged with the majority band, and their contribution to gap shrinkage is not included in the model.

The model outlined has two attractive features:

(i) As it does not compete with models based on many-body theory it cannot claim to be as fundamental as other models recently proposed. But the benefit arising from this is its great simplicity. It can be applied to over a range of temperatures and carrier concentrations with considerable ease.

(ii) It is essentially parameter-free. The only variables which enter are those which must be involved in any theory of band-gap shrinkage: effective masses, temperature, carrier concentrations and dielectric constant. (The latter would enter a many-body treatment in a different manner as a sum over electronic transitions.)

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant ECS-8203091 (A.N.), by Motorola, Inc., Mesa, AZ (F.A.L. and A.N.), and by Jet Propulsion Laboratory (JPL) Contract 956525. P.T.L. is indebted to Professor K.-F. Berggren for a discussion.

APPENDIX: CONSTRAINTS ON THE MODEL

In order not to introduce the parameter $\underline{a}$ into the band-gap shrinkage formula, it is necessary to satisfy (17). Now the error introduced in putting

$$e^{-\lambda a} \approx 1 - \lambda a$$

has the following values

λa	0.1	0.2	0.3
% error	0.6	2.4	5.9

Let us put the requirement

$$\lambda a < z \quad (z \sim 0.25)$$

on the model. We also have from (13) that

$$E_G(0,0) > e^2/\varepsilon a \quad .$$

It then follows that

$$\lambda < z/a < z\varepsilon\, E_G(o,o)/e^2 \quad .$$

Some algebra now shows, using the degenerate limit, that one requires

$$n^{1/6} < n_c^{1/6} \equiv \frac{z}{4}\, 10^{3/2} \left(\frac{\pi}{3}\right)^{1/6} \frac{(\varepsilon/10)^{3/2}}{(m_n/m)^{1/2}} \frac{E_G(0,0)}{I_o a_o^{1/2}} \quad .$$

Here I_o is 13.6 eV and a_o = 0.529 A is the Bohr radius. The critical concentration is

$$n_c = \frac{(\varepsilon/10)^9}{(m_n/m)^3} (4z)^6 E_G^{\,6} \times 0.659 \times 10^{20} \text{ cm}^{-3}$$

where E_G is expressed in eV. Using $\epsilon = 11.7$ and $E_G = 1.1$ eV, the constra' is for silicon

$$n < n_c = \frac{(4z)^6}{(m_n/m)^3} 4.80 \times 10^{20} \text{ cm}^{-3} \quad .$$

REFERENCES

1. A. Selloni and S. T. Pantelides, "Electronic structure and spectra of heavily doped n-type silicon," Phys. Rev. Lett., 49, 586 (1982).

2. K.-F. Berggren and B. E. Sernelius, "Band-gap narrowing in heavily doped many-valley semiconductors," Phys. Rev., B24, 1971 (1981).

3. G. D. Mahan, "Energy gap in Si and Ge: Impurity dependence," J. Appl. Phys., 51, 2634 (1980).
P. A. Sterne and J. C. Inkson, "Comment on energy gap in Si and Ge: Impurity dependence,: J. Appl. Phys., 52, 6432 (1981).

4. L. Engstron, "Band-gap narrowing in n-Si and n-Ge; effects of non-linear impurity scattering," J. Phys., C16, 1437 (1983).

5. W. J. Moore, Physical Chemistry (New York: Prentice-Hall, 1955), p. 463.

6. F. J. Morin and J. P. Maita, "Conductivity and Hall effect in the intrinsic range of germanium," Phys. Rev., 94, 1525 (1954).

7. J. C. Inkson, "The effect of electron interaction on the band gap of extrinsic semiconductors," J. Phys., C9, 1177 (1976).

8. R. A. Abram, G. J. Rees, B. L. H. Wilson, "Heavily doped semiconductors and devices," Advances in Physics, 27, 799 (1978).

9. C.-T. Sah, P. C. H. Chan, C.-K. Wang, R.L.-Y. Sah, K. A. Yamakawa and R. Lutwack, "Effect of zinc impurity on silicon solar-cell efficiency," IEEE Trans. Electron Devices, ED-28, 304 (1981).

10. V. L. Bonch-Bruevich, The Electronic Theory of Heavily Doped Semiconductors (New York: American Elsevier Co., 1966).

11. V. L. Bonch-Bruevich and R. Rozman, "On the theory of the absorption of light in heavily doped semiconductors," Soviet Physics Solid State, 6, 2016 (1965).

12. H. P. D. Lanyon and R. A. Tuft, "Bandgap narrowing in moderately to heavily doped silicon," IEEE Trans. Electron Devices, ED-26, 1014 (1979).

13. C. D. Thurmond, "The standard thermodynamic functions for the formation of electrons and holes in Ge, Si, GaAs and GaP", J. Electrochem. Soc., 122, 1133 (1972).

14. P. T. Landsberg, "Einstein and statistical thermodynamics III: The diffusion-mobility relation in semiconductors," Europ. J. Phys., 2, 213 (1981).

15. A. Neugroschel, S. C. Pao, and F. A. Lindholm, "A method for determining energy gap narrowing in highly doped semiconductors," IEEE Trans. Electron Devices, ED-29, 894 (1982).

16. H. Fritzsche, "The metal-non-metal transitions in disordered systems," Proceedings of the 19th Scottish Summer School in Physics, 1978.

17. J. W. Slotboom and H. C. DeGraaff, "Measurements of bandgap narrowing in Si bipolar transistors," Solid-State Electron., 19, 857 (1976).

18. D. D. Tang, "Heavy doping effects in p-n-p bipolar transistors," IEEE Trans. Electron Devices, ED-27, 563 (1980).

19. A. W. Wieder, "Emitter effects in shallow bipolar devices," IEEE Trans. on Electron Devices, ED-27, 1402 (1980).

20. G. E. Possin, M. S. Adler, and B. J. Baliga, "Measurement of the p-n product in heavily doped epitaxial emitters," IEEE Trans. Electron Devices, ED-31, 3 (1984).

21. R. P. Mertens, J. L. van Meerbergen, J. F. Nijs, and R. J. Van Overstraeten, "Measurement of the minority-carrier transport parameters in heavily doped silicon," IEEE Trans. on Electron Devices, ED-27, 949 (1980).

22. H. E. J. Wulms, "Base current of I^2L transistors," IEEE J. Solid-State Circuits, 12, 143 (1977).

23. A. Neugroschel and F. A. Lindholm, "Evidence for low diffusivity and mobility of minority carriers in highly doped Si and interpretations," Appl. Phys. Lett., 42, 176 (1983).

24. A. Neugroschel and F. A. Lindholm, "Reply on comments on 'A method for determining energy gap narrowing in highly doped semiconductors'", IEEE Trans. Electron Devices, ED-31, 124 (1984).

25. J. Wagner, "Photoluminescence and exitation spectroscopy in heavily doped n- and p-type silicon, Physical Review B, 29, 2002 (1984).

26. H. D. Barber, "Effective mass and intrinsic concentration in silicon," Solid-State Electronics, 10, 1039 (1967).

27. W. D. Dumke, "Comparison of band-gap shrinkage observed in luminescence from n^+-Si with that from transport and optical absorption measurements," Appl. Phys. Lett., 42, 196 (1983).

28. W. D. Dumke, "Band-gap narrowing from luminescence in p-type Si," J. Appl. Phys., 54, 3200 (1983).

29. F. L. Madarasz, J. E. Lang and P. M. Hemeger, "Effective masses for nonparabolic bands in p-type silicon," J. Appl. Phys., 52 4646 (1981).

30. M. Balkanski, A. Aziza and E. Amzallag, "Infrared absorption in heavily doped n-type Si," Phys. Status Solidi, 31, 323 (1969).

31. A. A. Volf'son and V. K. Subashiev, "Fundamental absorption edge of silicon heavily doped with donor or acceptor impurities," Sov. Phys.-Semicond., 1, 327 (1967).

32. P. E. Schmid, "Optical absorption in heavily-doped silicon," Phys. Rev., 23, 5531 (1981).

FIGURE CAPTIONS

Fig. 1 Schematic diagram showing the decomposition of the band-gap energy into W and the work done against attraction. More carriers are assumed present for curve 1 than for curve 2.

Fig. 2 A plot of R (equation (20)) against electron concentration n for Si. The values of R below the curve are forbidden in the present approximation, which begins to fail near $n \sim n_c = 4 \times 10^{20}$ cm^{-3}.

Fig. 3 Gap shrinkage ΔE_G as inferred from transport measurements for n-type layers from various sources at a mean temperature of ~340 K. The curves are based on equation (21), the upper curve is for $m^*/m = 1.45$ [10], the lower curve is for $m^*/m = 1.10$, and $\varepsilon = 11.7$ (Si) has also been used. The horizontal axis is the majority carrier concentration.

Fig. 4 Comparison of (21) with recent optical data at 5 K [21].

Fig. 5 Comparison of (21) with recent photoemission data at 300 K [23,24]. The data from [23,24] were corrected for degeneracy, as described in the text.

Fig. 6 Comparison of two theoretical curves for Si at 300 K based on equation (21). The lower curve applies to doped n-type silicon with $m_n/m = 1.45$ [10]. The upper curve applied to intrinsic excited silicon with $n = p$, $m_n/m = 1.45$ and $m_{p/m} = 1.08$. Also shown, for comparison, is a theoretical curve for n-type GaAs with $m_n/m = 0.066$ and $\varepsilon = 12.5$.

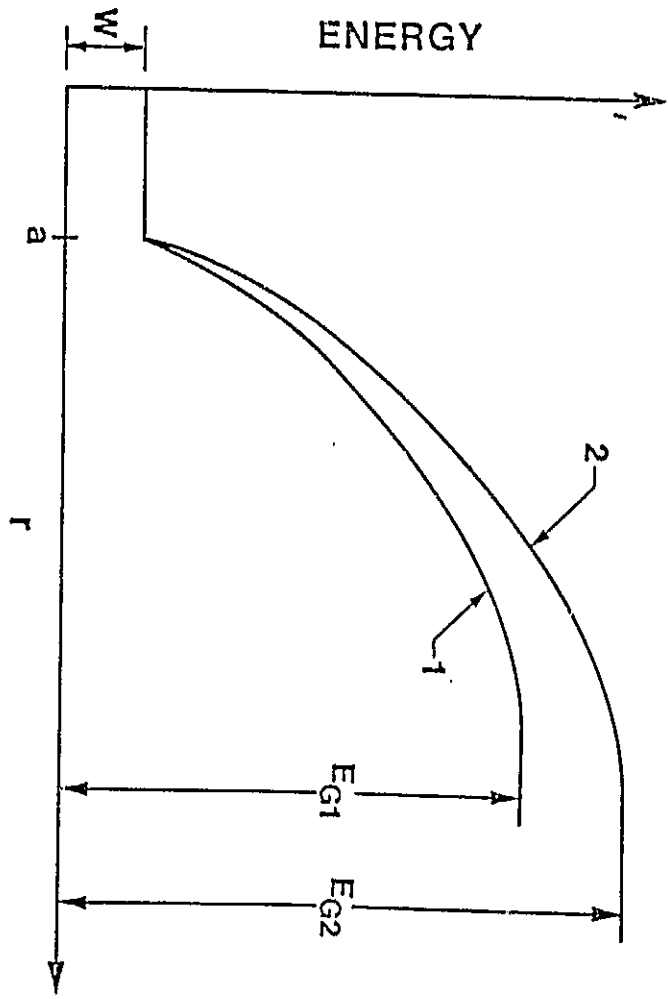


Fig. 1

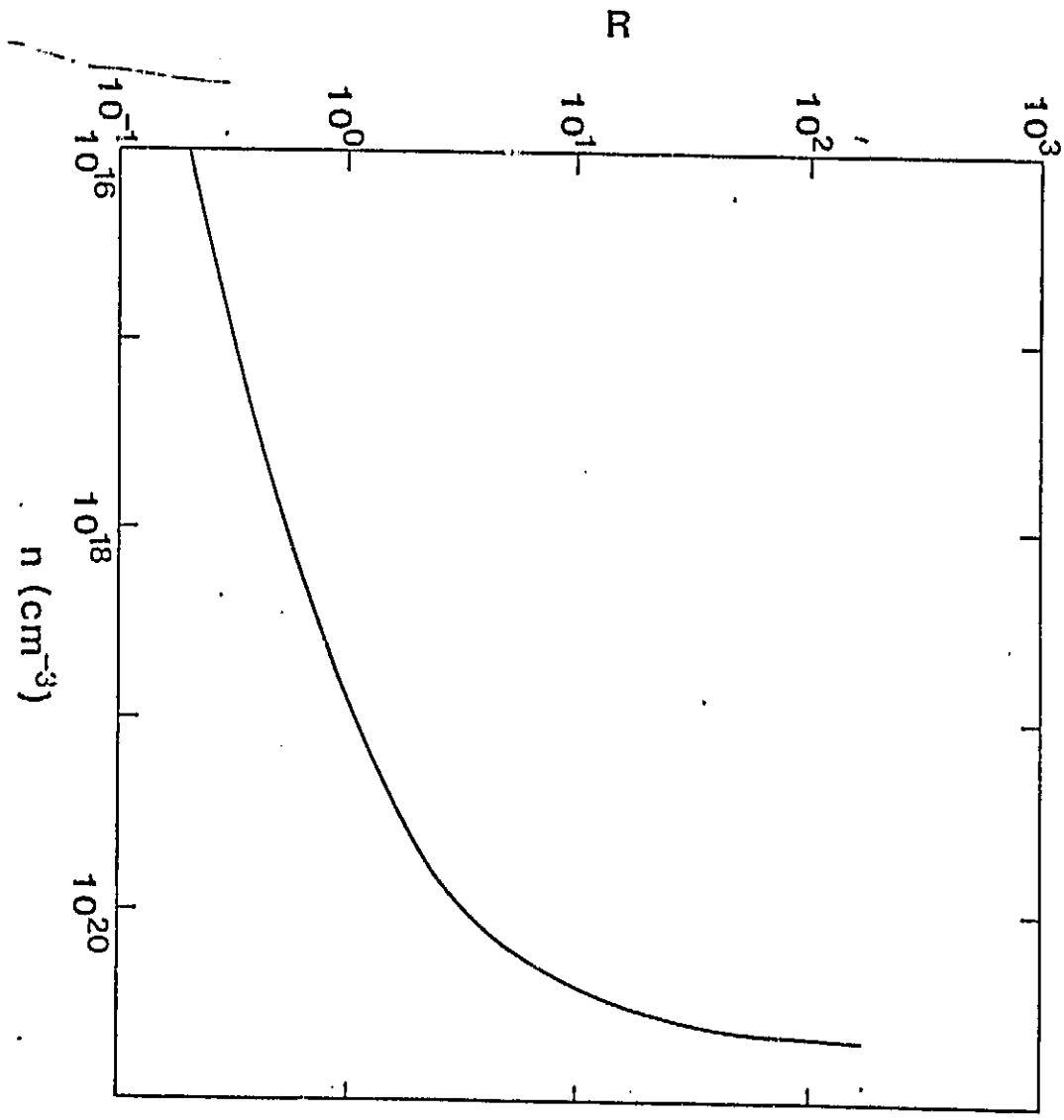


Fig. 2

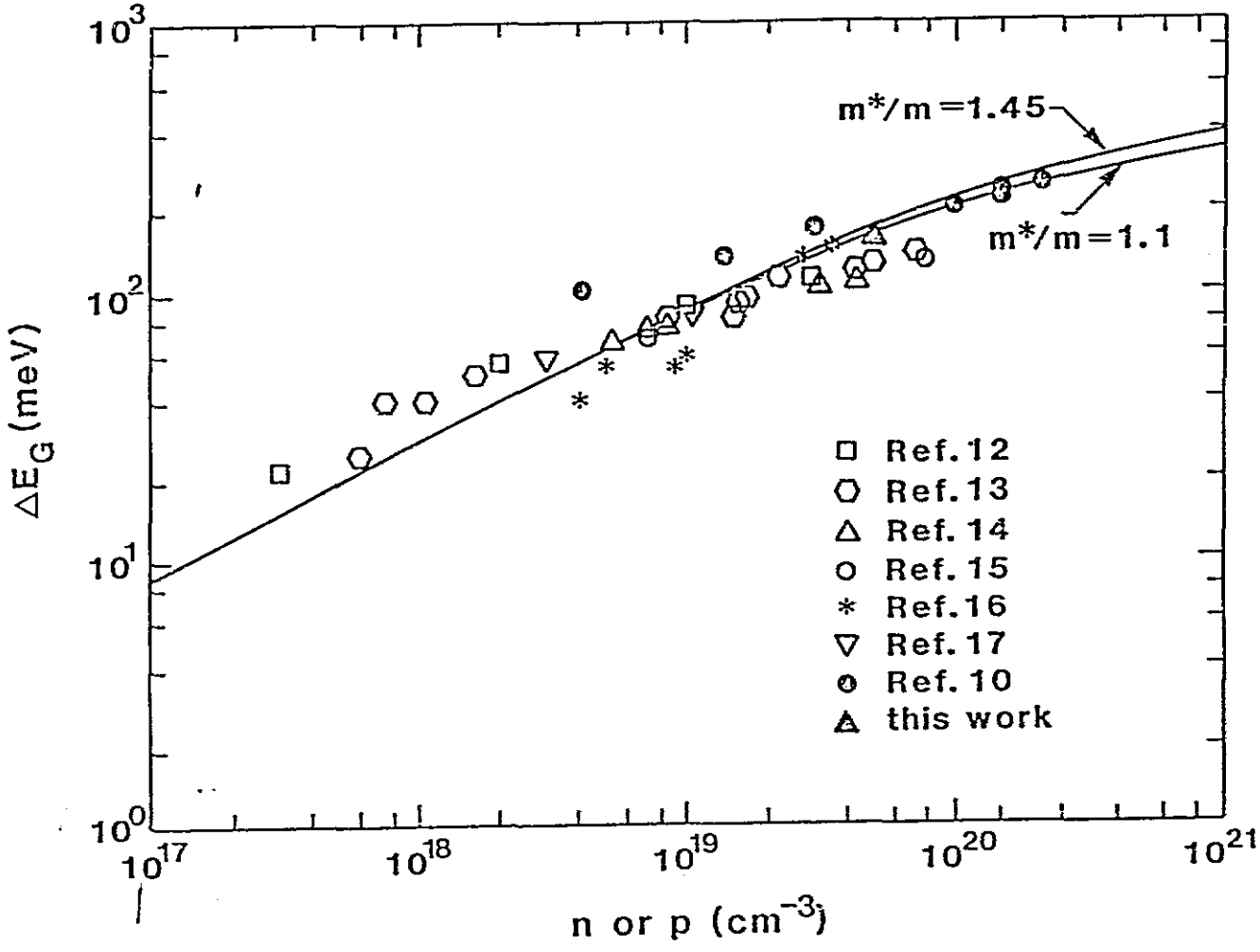

10^3
10^2
10^1
10^0
ΔE_G (meV)
10^{17}
10^{18}
10^{19}
10^{20}
10^{21}
n or p (cm^{-3})
m*/m=1.45
m*/m=1.1
Ref. 12
Ref. 13
Ref. 14
Ref. 15
Ref. 16
Ref. 17
Ref. 10
this work

Fig. 3

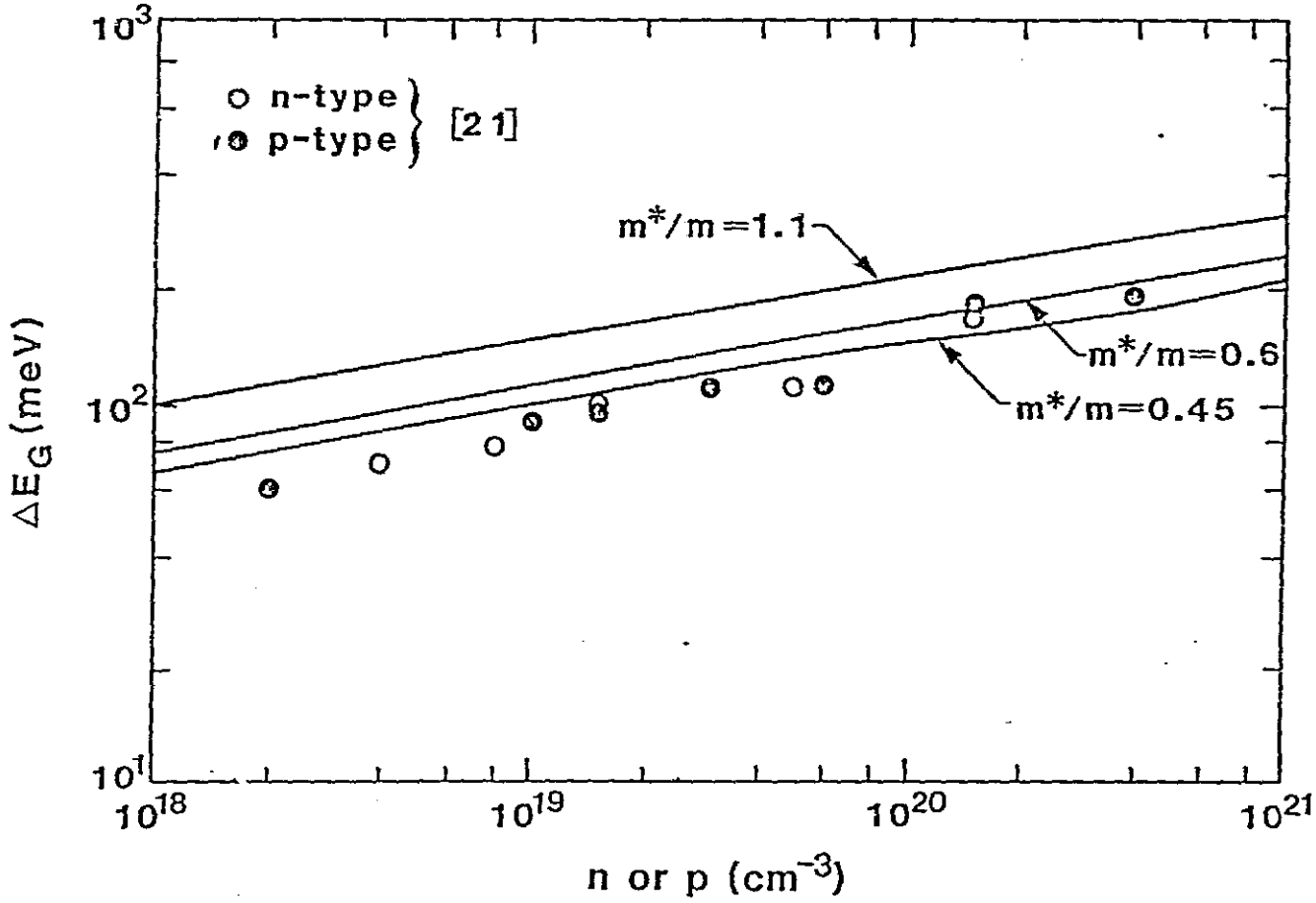

10^3
10^2
10^1
ΔE_G (meV)
o n-type
p-type
[21]
$m^*/m=1.1$
$m^*/m=0.6$
$m^*/m=0.45$
10^{18}
10^{19}
10^{20}
10^{21}
n or p (cm^{-3})

Fig. 4

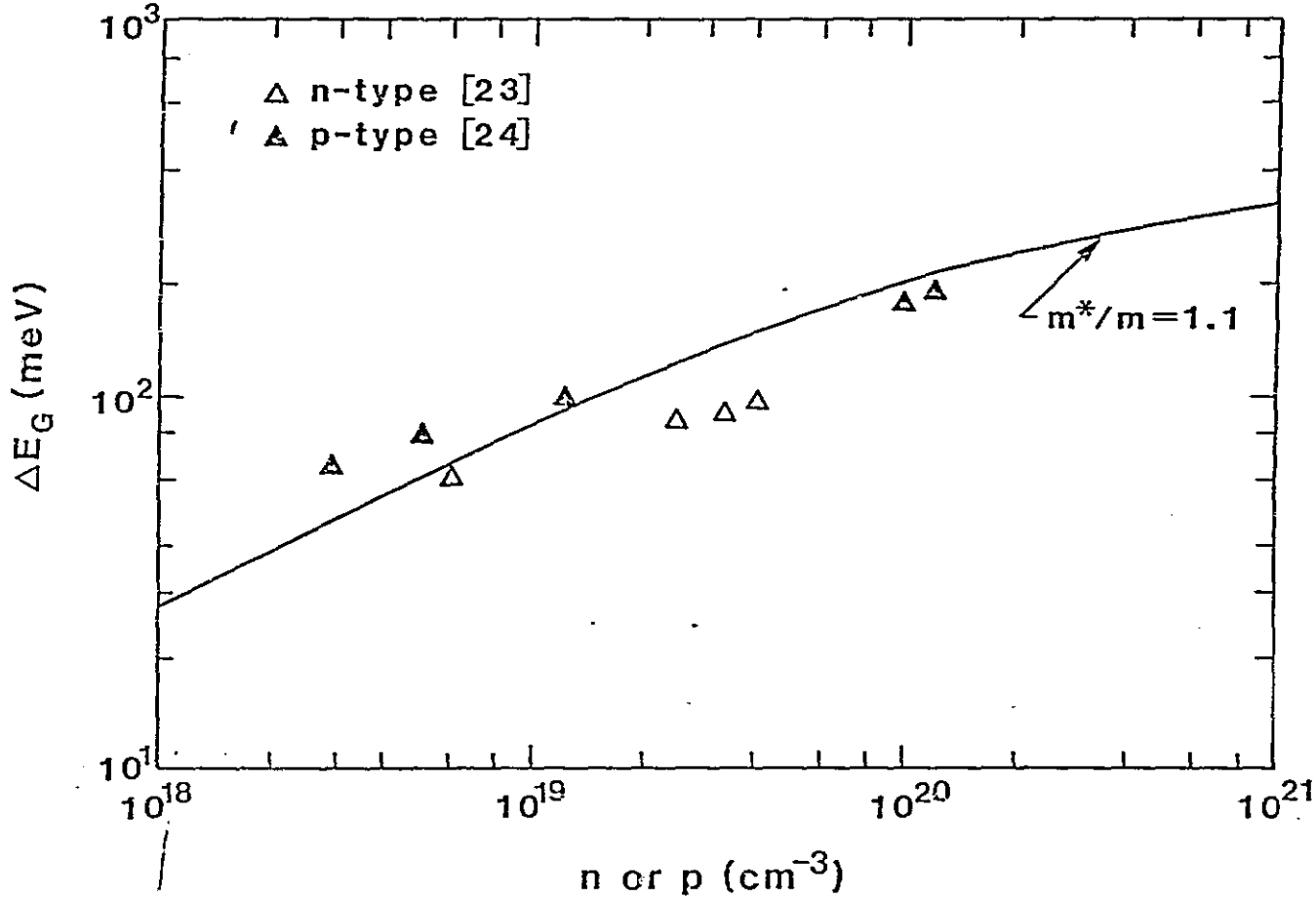

10^3
10^2
10^1
ΔE_G (meV)
△ n-type [23]
▲ p-type [24]
m*/m=1.1
10^{18}
10^{19}
10^{20}
10^{21}
n or p (cm^{-3})

Fig. 5

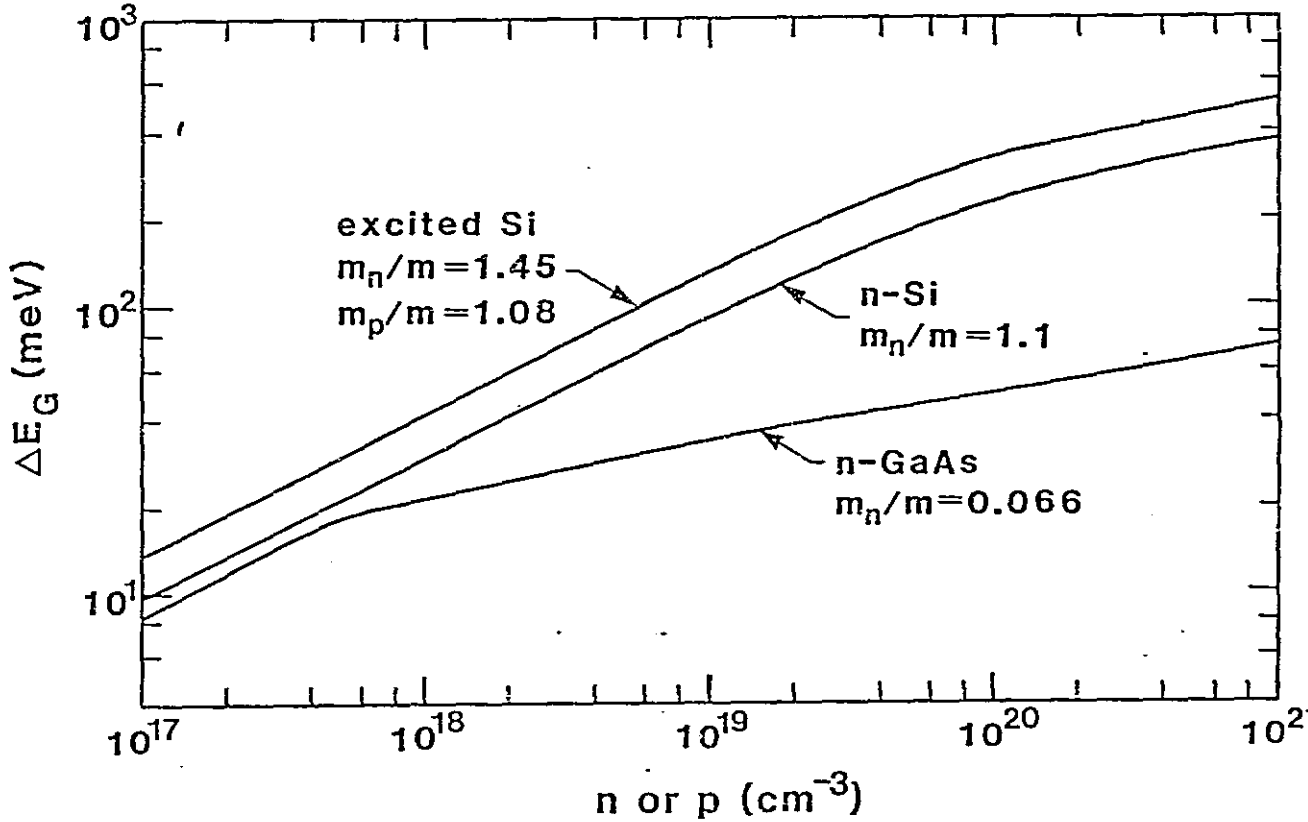

10^3
10^2
10^1
ΔE_G (meV)
excited Si
$m_n/m=1.45$
$m_p/m=1.08$
n-Si
$m_n/m=1.1$
n-GaAs
$m_n/m=0.066$
10^{17}
10^{18}
10^{19}
10^{20}
10^{21}
n or p (cm^{-3})

Fig. 6